

10/562096

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
13 January 2005 (13.01.2005)

PCT

(10) International Publication Number
WO 2005/003308 A3(51) International Patent Classification:
G01N 33/48 (2006.01)Sujoy [IN/US]; GlaxoSmithKline, Five Moore Drive,
PO Box 13398, Research Triangle Park, NC 27709 (US).
KUMAR, Vinod, D. [US/US]; GlaxoSmithKline, 709
Swedeland Road, King of Prussia, PA 19406 (US).(21) International Application Number:
PCT/US2004/019932

(22) International Filing Date: 22 June 2004 (22.06.2004)

(74) Agents: LEVY, David, J. et al.; GlaxoSmithKline, Five
Moore Drive, PO Box 13398, Research Triangle Park, NC
27709 (US).

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/482,420 25 June 2003 (25.06.2003) US(81) Designated States (*unless otherwise indicated, for every
kind of national protection available*): AE, AG, AL, AM,
AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN,
CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI,
GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE,
KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD,
MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG,
PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM,
TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM,
ZW.(71) Applicant (*for all designated States except US*):
SMITHKLINE BEECHAM CORPORATION
[US/US]; One Franklin Plaza, Philadelphia, PA 19101
(US).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): VAGARWAL,
Pankaj [IN/US]; GlaxoSmithKline, 709 Swedeland Road,
King of Prussia, PA 19406 (US). HURLE, Mark, Robert
[US/US]; GlaxoSmithKline, 709 Swedeland Road, King of
Prussia, PA 19406 (US). KABNICK, Karen, Stephanie
[US/US]; GlaxoSmithKline, 709 Swedeland Road, King
of Prussia, PA 19406 (US). LIU, Liwen [CN/US]; Glaxo-
SmithKline, Five Moore Drive, PO Box 13398, Research
Triangle Park, NC 27709 (US). MAGID-SLAV, Michal
[IL/US]; GlaxoSmithKline, 709 Swedeland Road, King of
Prussia, PA 19406 (US). MCALLISTER, Paul, Robert
[US/US]; GlaxoSmithKline, 709 Swedeland Road, King
of Prussia, PA 19406 (US). SEARLS, David, Burdette
[US/US]; GlaxoSmithKline, 709 Swedeland, King of Prus-
sia, PA 19406 (US). TATSUOKA, Kay, Satoshi [US/US];
GlaxoSmithKline, 709 Swedeland Road, King of Prussia,
PA 19406 (US). ZAYKIN, Dmitri, V [RU/US]; Glaxo-
SmithKline, Five Moore Drive, PO Box 13398, Research
Triangle Park, NC 27709 (US). REISDORF, William,
Charles, Jr. [US/US]; GlaxoSmithKline, 709 Swede-
land Road, King of Prussia, PA 19406 (US). GHOSH,(84) Designated States (*unless otherwise indicated, for every
kind of regional protection available*): ARIPO (BW, GH,
GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM,
ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI,
FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE,
SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN,
GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— with international search report
— before the expiration of the time limit for amending the
claims and to be republished in the event of receipt of
amendments(88) Date of publication of the international search report:
31 August 2006For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.

(54) Title: BIOLOGICAL DATA SET COMPARISON METHOD

(57) Abstract: A method of identifying a relationship between a set of one or more candidate biomolecules and a set of one or more reference biomolecules, the method including inputting to a computer a query set describing the one or more candidate biomolecules; comparing the query set with a target database describing the one or more reference biomolecules wherein the one or more reference biomolecules grouped into one or more buckets and wherein the one or more reference biomolecules of each bucket share a common property; counting a number of matches between each query set and each buckets of the target database; and statistically analyzing the number of matches to each bucket wherein the presence of a statistically significant match identifies a relationship between a the query set and a bucket of the target database.

WO 2005/003308 A3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US04/19932

A. CLASSIFICATION OF SUBJECT MATTER IPC: G01N 33/48(2006.01) USPC: 702/19 According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) U.S. : 702/19 Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) Please See Continuation Sheet		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	LEIBOWITZ et al. "MUSTA - A General, Efficient, Automated Method for Multiple	1-31, 33-40
---	Structure Alignment and Detection of Common Motifs: Application to Proteins," Journal of	32, 41
Y	Computational Biology (2001) Volume 8, Number 2, pages 93-121, especially abstract, pages	
	95-96, 100-104.	
Y	YAP et al. "Parallel Homologous Sequence Searching in Large Database," Fifth Symposium	1-41
	on the Frontiers of Massively Parallel Computation, 1995. Proceedings. 'Frontiers '95'.	
	February 1995, page 231-237 especially pages 231-233.	
Y	YAP et al. "Parallel Computation in Biological Sequence Analysis," IEEE Transactions on	1-41
	Parallel and Distributed Systems (March 1998) Volume 9, Number 3, pages 283-294	
	especially pages 284-292.	
A	US 5,799,312 (RIGOUTSOS) 25 August 1998 (25.08.1998), columns 1-4	1-41
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.		
* Special categories of cited documents:		
"A"	document defining the general state of the art which is not considered to be of particular relevance	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"E"	earlier application or patent published on or after the international filing date	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O"	document referring to an oral disclosure, use, exhibition or other means	"&" document member of the same patent family
"P"	document published prior to the international filing date but later than the priority date claimed	
Date of the actual completion of the international search 17 May 2006 (17.05.2006)		Date of mailing of the international search report 06 JUL 2006
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450 Alexandria, Virginia 22313-1450 Facsimile No. (571) 273-3201		Authorized officer Jerry Lin Telephone No. (571) 272 1600

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US04/19932

Box No. II Observations where certain claims were found unsearchable (Continuation of Item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
Please See Continuation Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of any additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-41

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US04/19932

BOX III. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group 1, claim(s) 1-41, drawn to a method of identifying the relationship between a biomolecule and a reference biomolecule.

Group 2, claim(s) 42-54, drawn to a data structure with three fields.

Group 3, claim(s) 55, drawn to a method of making a target database.

The inventions listed as Groups 1-3 do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Group 1 has the special technical feature of comparing biomolecules to reference biomolecules. Group 2 has the special technical feature of a data structure with three fields. Group 3 has the special technical feature of making a database. Each of the Groups do not share the same special technical feature with the other groups, thus Groups 1-3 do not relate to a single general inventive concept.

Continuation of B. FIELDS SEARCHED Item 3:

West, Medline, embase, biosis

Terms: query, protein, DNA, RNA, bucket